

WHAT IS CLAIMED IS:

1. A computerized method for determining whether a biological sequence has certain characteristic comprising:
obtaining a plurality of evidence about the characteristic, wherein at least one evidence is sequence annotation; and
determining the characteristic using a Bayesian analysis of the evidence.
2. The method of Claim 1 wherein the step of determining comprises:
defining the prior probability of the biological sequence having the characteristic;
estimating the probability of the evidence assuming the hypothesis is true; and
calculating the probability that the hypothesis is true.
3. The method of Claim 2 wherein the step of calculating is performed according to Bayes' rule.
4. The method of Claim 3 wherein the biological sequence is a nucleic acid sequence and the characteristic is the orientation of the biological sequence.
5. The method of Claim 4 wherein the nucleic acid sequence represents a cluster of nucleic acid sequences including at least one EST sequence.
6. The method of Claim 5 wherein the plurality of evidence comprises evidence from poly-A/T tail analysis, inferred splice sites; and external sequence annotation.
7. The method of Claim 6 wherein the external sequence annotation comprises RNA label and EST label.

8. The method of Claim 7 further comprising testing a null hypothesis that the orientation determination is correct and conflicting evidence observed is due to random error.
9. A computerized method for designing nucleic acid probe arrays comprising:
obtaining a plurality of evidence about at least one characteristic of a target nucleic acid sequence, wherein at least one evidence is sequence annotation;
determining the characteristic using a Bayesian analysis of the evidence; and
defining a target region based upon the characteristic; and
selecting probes against the target region.
10. The method of Claim 9 wherein the step of determining comprises defining the prior probability that a hypothesis that the target nucleic acid sequence has the characteristic; estimating the probability of the evidence assuming the hypothesis is true; and
calculating the probability that the hypothesis is true.
11. The method of Claim 10 wherein the step of calculating is performed according to Bayes' Rule.
12. The method of Claim 11 wherein the characteristic is the orientation of the target nucleic acid sequence.
13. The method of Claim 12 wherein the nucleic acid sequence represents a cluster of nucleic acid sequences including at least one EST sequence.
14. The method of Claim 13 wherein the plurality of evidence comprises evidence from poly-A/T tail analysis, inferred splice sites; and external sequence annotation.

15. The method of Claim 14 wherein the external sequence annotation comprises RNA label and EST label.
16. The method of Claim 15 further comprising testing a null hypothesis that the orientation determination is correct and conflicting evidence observed is due to random error.
17. A system for characterizing a biological sequence comprising a processor; and a memory coupled with the processor, the memory storing a plurality of machine instructions that cause the processor to perform logical steps of :
obtaining a plurality of evidence about the characteristic, wherein at least one evidence is sequence annotation; and
determining the characteristic using a Bayesian analysis of the evidence.
18. The system of Claim 17 wherein the step of determining comprises:
defining the prior probability of the biological sequence having the characteristic;
estimating the probability of the evidence assuming the hypothesis is true; and
calculating the probability that the hypothesis is true.
19. The system of Claim 18 wherein the step of calculating is performed according to Bayes' Rule.
20. The system of Claim 19 wherein the biological sequence is a nucleic acid sequence and the characteristic is the orientation of the biological sequence.
21. The system of Claim 20 wherein the nucleic acid sequence represents a cluster of nucleic acid sequences including at least one EST sequence.
22. The system of Claim 22 wherein the plurality of evidence comprises evidence from poly-A/T tail analysis, inferred splice sites; and external sequence annotation.

23. The system of Claim 22 wherein the external sequence annotation comprises RNA label and EST label.
24. The system of claim 23 further comprising testing a null hypothesis that the orientation determination is correct and conflicting evidence observed is due to random error.
25. A system for characterizing a biological sequence comprises a processor; and a memory coupled with the processor, the memory storing a plurality of machine instructions that cause the processor to perform logical steps of :
obtaining a plurality of evidence about at least one characteristic of a target nucleic acid sequence, wherein at least one evidence is sequence annotation;
determining the characteristic using a Bayesian analysis of the evidence; and
defining a target region based upon the characteristic; and
selecting probes against the target region.
26. The system of Claim 25 wherein the step of determining comprises defining the prior probability that a hypothesis that the target nucleic acid sequence has the characteristic; estimating the probability of the evidence assuming the hypothesis is true; and
calculating the probability that the hypothesis is true.
27. The system of Claim 26 wherein the step of calculating is performed according to Bayes' Rule.
28. The system of Claim 27 wherein the characteristic is the orientation of the target nucleic acid sequence.

29. The system of Claim 28 wherein the nucleic acid sequence represents a cluster of nucleic acid sequences including at least one EST sequence.
30. The system of Claim 29 wherein the plurality of evidence comprises evidence from poly-A/T tail analysis, inferred splice sites; and external sequence annotation.
31. The system of Claim 30 wherein the external sequence annotation comprises RNA label and EST label.
32. The system of Claim 31 further comprising testing a null hypothesis that the orientation determination is correct and conflicting evidence observed is due to random error.